Notes e-book

**Content overall:**

* Generic chapters (all blocks, needs to be done now)
* Data analysis methods theory (all blocks, some needs to be done now, some can wait e.g. PLS)
* MST chapters (block 3, missing)
* F&MCR chapters (block 4, missing)
* TFIH chapters (block 1, needs to be done now)

General formatting: Whenever in narrative mentioning packages put them in bold: \*\*devtools\*\*

**To do list** is found after the Index...

**Index** **(work in progress):**

1. Introduction to the book
2. (#, 01) Introduction to R – generic DONE
   1. (##) How to get started - understanding R (and RStudio) DONE
      1. (###, 01) Organise and save scripts DONE
   2. (##) How to import data DONE (there is no text, so easy 😊)
      1. (###) Import data from R-package DONE
      2. (###) Importing a csv file DONE
      3. (###) Importing an Excel file/sheet DONE
      4. (###) Clipboard import DONE
      5. (###) Looking at the imported elements DONE
      6. (###) Numbers and factors - changing categorisation DONE
   3. (##, 01) How to merge two datasets
      1. (###, 01) Import and merge in R (DONE)
         1. (####, 01) Edit using Tidyverse DONE
         2. (####, 01) Merging datasets DONE
            1. (#####, 01) Adding survey to buffet
            2. (#####, 01) Adding buffet to survey (DONE)
   4. (##, 01) How to save the data (semi- DONE – check if the explanation is sufficient now
   5. (##, 01) How to export data / results DONE
   6. (##, 01) Ready for analysis DONE
3. (#, 02) Libraries (semi- DONE – check if text is sufficient
4. (#, 03) Descriptive statistics
   1. ## Descriptives for a continuous variable DONE
      1. ### Mean/ median DONE
      2. ### Variance DONE
      3. ### Standard deviation DONE
      4. ### Calculations DONE
   2. ## Distributions of count data (MANGLER: skal vi overveje at lave dette med CATA data i stedet for? jeg er ikke helt vild med legumes datasættet nemlig.... men kan også bare blive stående)
   3. ## Aggregate DONE
   4. ## Tidyverse DONE
5. Inferential statistics Morten tjekke hele dette kapitel.
   1. Intro
   2. Hypothesis testing
      1. T-test
      2. F-test
   3. Confidence intervals
   4. ANOVA - DONE
      1. Post hoc test: Tukey’s Honest Significant Difference (HSD) DONE
6. (#, 05) Plotting data DONE
   1. ## Histograms and boxplots semi- DONE – check if text is sufficient
   2. ## Scatter plots (semi- DONE – check if text is sufficient
   3. ## How to export plots DONE
7. (#, 06) Introduction to PCA and multivariate data (ved ikke om der mangler noget???)
   1. ## Interpreting model output (MANGLER: Eksempel på PCA output? Jeg foreslår PCA på sensorik datasæt lagt ind I Excel filen: BeefSensoryProfile. Jeg har beskrevet datasættet her)
8. (#, 07) Introduction linear and mixed models DONE
   1. ## What is a linear model?
   2. ## Normal and Mixed models DONE
      1. ### Normal model DONE
      2. ### Mixed model DONE
9. # Buffet data
   1. ## Importing and looking at the buffet data (including descriptive stats)
10. #Simple survey data – laves med specifikke data for dette kursus og henvise til
    1. Henvise til kap ”5” om inference/ Inferential statistics
    2. Mixed linear models på buffet data
       1. Simple models
       2. Multivariable models
    3. PCA?
11. # Buffet and survey data combined
    1. ## Merging data sets
    2. ## ANALYSIS??? Jeg ved ikke hvad der kan laves her
12. MST Exercises – spørge Morten om hvad der skal med
13. FMCR COURSE HERE
    1. Large survey data
    2. Table one
    3. Cluster analysis
       1. K means
       2. Choosing segments by attributes
    4. Cluster profiling
       1. Logistic regression
14. FMCR exerises on big data EU
15. # CATA data (Check-all-that-apply)
    1. ## Importing and looking at the beer data (MANGLER: checke om data er korrekt)
    2. ## Two versions of the data (MANGLER: punktopstilling. put in picture. Is my interpretation of the functions correct? Forklare de to do outputs. Indsætte kapitel link)
    3. ## Cochran’s Q test
       1. ### Post hoc contrasts (MANGLER: Morten: explain the code in words. Skal man bruge p.value eller p.adjust?)
       2. ### For all attributes in one run (nice to know) (MANGLER: Morten explain the code in words)
    4. ## PCA on CATA data (MANGLER: explain the code in words, BOM: interpretation of the PCA plot – send by email. Bodil tilføjer når det kommer)
16. # Hedonic rating scores (DONE)
    1. (##, 10) Plotting liking scores (MANGLER: fortolkning af output, Bom)
    2. (##, 10) Simple mixed models (MANGLER: fortolkning af output, Bom)
       1. (###, 10) Post hoc test (MANGLER: stor note er slettet, det skal udbygges med tiden?)
    3. (##, 10) Multivariable models (DONE)
       1. ### Additive models (DONE)
       2. ###Effect modification and Interactions (DONE)
17. # CATA and hedonics (DONE)
    1. ## Individual attributes and liking (DONE)
       1. ### An example with Refreshing (DONE)
       2. ### All attributes (MANGLER: explanation for models, is this a “nice to know”??)
       3. ### A beer centric model (MANGLER: explanation for models)
    2. ## PCA on CATA and Liking (MANGLER: explanation for models)
18. # Projective mapping (MANGLER: ALT, men vi ser ikke på det i år 2022)
    1. PCA on projective mapping data
    2. (MFA on projective mapping data)
19. # TFIH exercises (MANGLER: ALT, der er skrevet et skelet dog)
20. # Preference mapping (først til 2023: PLS on CATA and liking data)

**TO DO LIST:**

**EDITS to version 0.0 – missing in present version:**

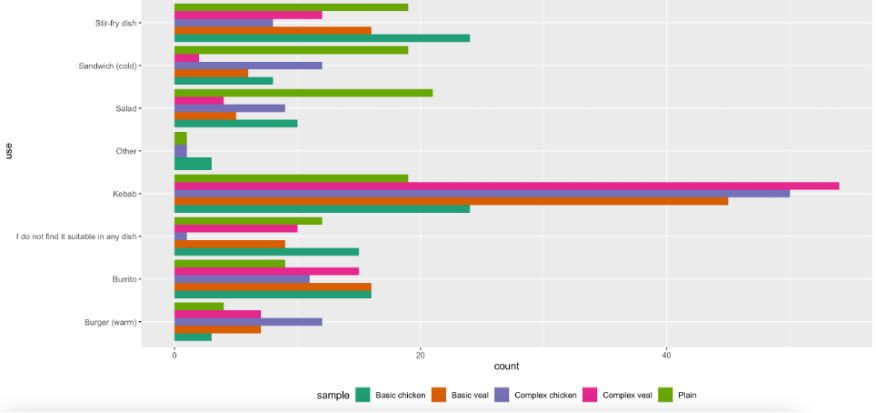
1. **Chapter 2. introduction to R, 2.2.3 Importing an Excel file/sheet:** In the following code, Is 'BuffetSurvey' supposed to be connected to the sheet SurveyData in excel? Are the name of the sheet in the code and in the excel file not matching?

library(readxl)  
BuffetConsumption <- read\_excel('./data/DatasetRbook.xlsx',sheet = 'BuffetData')  
BuffetSurvey <- read\_excel('./data/DatasetRbook.xlsx',sheet = 'BuffetSurvey')

1. **Chapter 8.4 PCA on CATA**: add explanation to the [] in the PCA model sentences. Morten did in class describe “#hard brackets is to make a subset of the [1-10,3:29] means first ten rows, and columns 3-29, -1 -means I do not want the first column”
2. **Chapter 8.4 PCA on CATA**: add more info in model: varname.size = 8, and groups (to colour by column beer and elipses) as Morten did in class, also how to change size of sample names
3. **Chapter 9.1 Plotting liking scores:** add explanation for model
4. **General in PCA plots**: hvordan man ændrer på akserne på PCA plottet, så alle variablene kan læses og ikke skæres af. Hvordan ændrer man størrelsen på scores (ikke loadings/variable)? Morten:

m <- prcomp(beercata[,3:29], scale. = T)  
ggbiplot::ggbiplot(m, groups = beercata$Beer,varname.size = 2, ellipse =T) +   
geom\_point(size = 2, aes(color = beercata$Beer)) +   
  ylim(c(-10,10)) + xlim(c(-10,10))

1. **Chapter 8.3.1 CATA, post hoc test**: can we find a solution to the letter based representation here too? (as with linear models)
2. **Chapter 9.2.1 Post hoc test (linar models):** Switch the two pairwise comparison methods. I would like the Tukey letters to be the first, as that is the one, we use the most.
3. **Add chilli and pasta datasets** to Excel sheet ”DatasetRbook” in Dropbox. – is this done?
4. Hvordan **ændrer man rækkefælge på sine samples i plots** ( se vedhæftet kode + nedenstående plot) dette giver samples i vilkårlig rækkefælge, kan man ændre denne, sådan at plain eks kommer først?



Morten svarer:

prøv at factor(,levels = c('lev1','lev2',...))

hvor i smider lev1 ogg lev2,.. til den relevante rækkefølge

og i størrelser, så se

theme(axis.text.y = element\_text())

også prøv at googgle theme() detaljer

1. **Chapter 8.3.1 Post hoc test** maybe what is the desired alpha level (some examples)
2. **Chapter 9.2 Simple mixed models**: last sentences (it seems liking... maybe just a repetition why?
3. **Chapter 10.1.2 All attributes:** add explanation to model, by now we do not know where all the different names and variables come from.
4. **Chapter 6**, first text bit is missing "variable" in the end: ...but mostly PCA is used for analysis of continuous (to be added: variable).
5. **Chapter 7.1** "Linear models are on of the most used statistical methods". - On = one.
6. **Chapter 7.1**, first text bit, last sentence: "then both of the models below is linear models". - Is = are.
7. **Chapter 8.3**, last text bit: "For this we need at post hoc test. - at = a.
8. **Chapter** **8.4**. First text bit: "For an introduction PCA" - missing a "to"?
9. **Chapter 9.2.1** Last text bit: "The resulst indicate Brown Ale and Ravnsborg Red is the most likable beers". - results spelled wrong.
10. **Chapter 9.2.1**. Same last text bit: "These three at the bottom is also not significantly different." - is =are
11. **Chapter 9.3.1** This text: "In the beer dataset we would like to know which of the explanatory variables that is most related to the liking." - "is" should be "are" I guess?
12. **Chapter 9.3.2**. Last sentence: "You cannot remove a single variable if an interaction including it is significant." - including should be included? then I understand it
13. **Chapter 10.1.2**.: after the barplots "For each attribute a interaction model will be used to qualify further analysis. - a = an.

**Bodil**

* Edit Word file to delete the stuff in the list, which has been fixed
* Simplify **FMCR** data for exercises, names of variables e.g.
  + Descr. stats
  + k means + own choices
  + Naming clusters
  + log reg on clusters incl. multiway (?)

**Morten**

* Add Beef data (from Sensory Science) to datapackage
* Add ryebread data (from Sensory Science) to datapackage (Tjek samtykke).
* Sæt beef m PCA som intro to PCA.
* Unwrap liking long and liking wide in chp 8.
* Change headings in chapter 10 to analysis of CATA + hedonic rating – subchapters ## PCA and ## PLS.
* Add more on mixed model.
* Code for exporting model output as an Excel table (make headline so it is easy to find)
* Check all Rmds for [notes to answer] – a lot, sorry
* Write a “Smart tips and tricks” section (all the short codes, you always forget), e.g. :
  + rm(list = ls())
  + str()
  + find more... 😊
  + ask Julius for more

**Bom**

* Explain projective mapping PCA output in e.g. screen cast
* Explain mixed models output in e.g. screen cast

**Helene**

* Intro – Chose what to do.
* Initiere basic plots (outliere and distributions)
  + Histogram
  + Boxplots
  + Point plots med liner til dommer
* Kontakte følgende ift. nye datasæt (se IFRO møde noter):
  + Thomas Bøker Lund, fødevareusikkerhed som hustandssurvey?
  + Sinne Smed, Indkøbsdata via GFK
  + Jørgen Dejgaard, kombinering af 4-5 forskellige databaser.
  + Lotte Holm & Thomas Bøker Lund: Mad og spisning i de nordiske lande. Forløbsdata. Tilbage fra 1990’erne.
  + COOP data, spørger til Aftagerpanelmøde
  + DTU kostundersøgelser
  + Varefakta

**Notes from meetings**

**Notes from meeting 3/10-2022 (Hennnnrik, Bodil, Helene, Morten)**

Make introduction chapter with

* Definition of variables (ordinal, conti,…)
* How to choose statistical model (pointing towards the chapters, what is a multivariate problem? What is a univariate problem? )

Make chapter on sensory = “Aroma summer course”

MST course (mostly just descriptives and plots and **adding linear models**)

Food+Meal CR course (plotting /descrp. K.means, logit, profiling)

In the Master in Technology, we will use the material as go-to for self-brush-up on stats.

Can this chapter on PCA be used as general introduction to multivariate / PCA for foreing students? [**Ask Åsmund to review**].

**Other ideas for data sets**

* Analysis of IFRO-kind-data (see meeting notes)
* House-hold data
* Single person
* New data: twitter, recepies, fødevarestyrelsen’s database…
* Marianne Thomsen?
* Dairy data?
* FQMC data? Nils?
* SOL Helene
* COOP dataset?
* App data? Maybe Inge Tetens?
* DTU kostundersøgelser – Helene kan nogle navne.

**Notes from IFRO meeting with Kia 24/10-2022:**

* Kia: Introducing Chi2 test in Excel in Social Science Methods, mentions R will come later in block 3 (us, MST course). Bør forstå bivariate sammenhænge efter hendes kursus.
* Kia kontakter Arne Hemmingsen ift hans R bog.
* Thomas data, fødevareusikkerhed som hustandssurvey? Helene kontakter.
* Indkøbsdata – Sinne Smed lektor <https://ifro.ku.dk/medarbejdere/?pure=da/persons/227352> ? Data ejes af firma. Tutorial datasæt kan måske frikøbes. Specialer. POCS. Ellers Sigrid Denver (???). Hedder GFK data. Helene kontakter.
* Jørgen Dejgaard data? Kombinering af 4-5 forskellige databaser. Helene kontakter.
* Lotte Holm & Thomas: Mad og spisning i de nordiske lande. Forløbs data. Tilbage fra 1990’erne. Helene kontakter.